

TNO PUBLIEK

Princetonlaan 6
3584 CB Utrecht
P.O. Box 80015
3508 TA Utrecht
The Netherlands

www.tno.nl

T +31 88 866 42 56
F +31 88 866 44 75

TNO report

TNO2021 R12442 | Final report

KEM9: Cumulative propagation effect of Groningen risk model component uncertainties on hazard and risk predictions

Date 9 September 2022

Author(s)

Copy no

No. of copies

Number of pages 32 (incl. appendices)

Number of
appendices

Sponsor Ministry of Economic Affairs and Climate Policy

Project name EZK Werkprogramma 2022

Project number 060.51941/01.07.05

All rights reserved.

No part of this publication may be reproduced and/or published by print, photoprint, microfilm or any other means without the previous written consent of TNO.

In case this report was drafted on instructions, the rights and obligations of contracting parties are subject to either the General Terms and Conditions for commissions to TNO, or the relevant agreement concluded between the contracting parties. Submitting the report for inspection to parties who have a direct interest is permitted.

© 2022 TNO

TNO PUBLIEK

Summary

The Groningen Model Chain's sensitivity on a number of model settings was tested. This was done by changing one model setting (model choice / assumption / parameter / input data) per sensitivity test with respect to a base case. The averaged risk of the test case is then compared with the base case. Most model setting variations result in a higher mean risk relative to the base case. We identified a small number of model settings that yield risk variations exceeding a 'significance' threshold of a factor 1.4 times the average risk of the base case. This threshold represents aleatory variability (irreducible uncertainty), based on the effects of alternative choices in the earthquake completeness magnitude M_{min} . The model settings exceeding this threshold are: the inclusion of period-2-period correlation in the site response of the ground motion model (GMM), alternative magnitude-frequency (MF) models, adapted fragility parameters of unreinforced masonry (URM) in the fragility model, and the use of the Coulomb stress in the calibration of the seismic source model. Most of these 'alternative model settings' were previously identified as having substantial impact on the risk results. This study confirms these findings.

Contents

	Summary	2
1	Introduction.....	4
2	Approach	7
2.1	Risk metrics	7
2.2	Seismic source model (SSM)	8
2.3	Ground motion model (GMM).....	13
2.4	Fragility and consequence model (FCM).....	13
2.5	Exposure database (ED)	15
3	Results	16
3.1	Overview	16
3.2	SSM: Calibration date range	19
3.3	SSM: Calibration range of included magnitudes	19
3.4	SSM: Magnitude-Frequency models	21
3.5	SSM: On-fault Coulomb stress vs. smoothed Coulomb stress	21
3.6	Alternative risk measure	22
3.7	Dissection of risk measure: mean risk, P10 and P90, and LPR of individual buildings.....	23
4	Discussion & conclusion	27
5	References	30
	Signature.....	32

1 Introduction

Hazard and risk predictions from induced earthquakes originating in and around the Groningen gas field as a result of gas extraction are simulated with a series of models that comprise the TNO Model Chain Groningen [TNO, 2020a]. This model chain is used to produce the yearly “public Seismic Hazard and Risk Analysis” (pSHRA) for the Groningen gas field.

In this report, we quantify the effect (and relative contribution) on the predicted hazards and risks resulting from epistemic uncertainty in model setting choices. Model settings are (1) selection of model chain components, (2) selection of input data and parameter values, and (3) model implementation choices and assumptions. Each component of the model chain passes on its results to the next component in the chain (e.g., seismic source model to ground motion model). Therefore, each choice of model settings is also propagated through the remainder of the model chain, changing risk and hazard relative to an alternative value/ implementation/ assumption for that model setting. Alternative choices for six model settings (with their respective weights) are incorporated in the pSHRA through the so-called logic tree (Figure 1). The sensitivity of the model chain for each logic tree branch is given by the variation in predicted risk (for example, Figure 2 for the pSHRA 2021).

A sensitivity analysis of three model settings outside of the logic tree has been reported with the pSHRA 2021 results [TNO, 2021a]. Figure 3 illustrates the need for a more thorough sensitivity analysis of model settings in the model chain. The results of this analysis are presented in this report.

The results of the extensive sensitivity analysis can be used for prioritization for further research within the framework of the public SHRA Groningen model development. There will be a particular focus on improved dealing with uncertainties that have a significant contribution (i.e., exceeding aleatoric variation in input data) to the Groningen pSHRA. Recommendations for the test cases that exceed this significance threshold are also presented in this report.

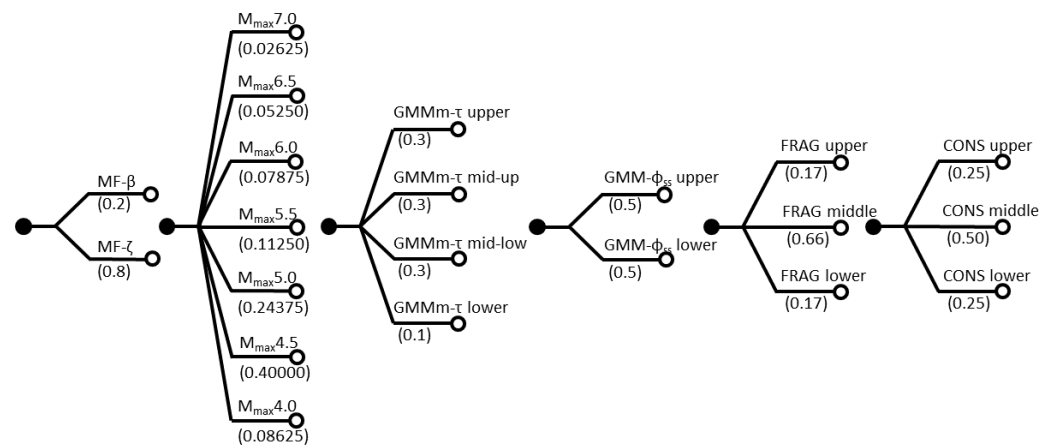


Figure 1: Current logic tree branches (including weights in brackets): Magnitude-Frequency distributions (MF); Maximum magnitude (M_{max}); median Ground Motion (GMMm- τ); Ground motion variability (GMM- Φ_{ss}); Fragility model (FRAG); Consequence model (CONS). Figure modified from [TNO, 2021a].

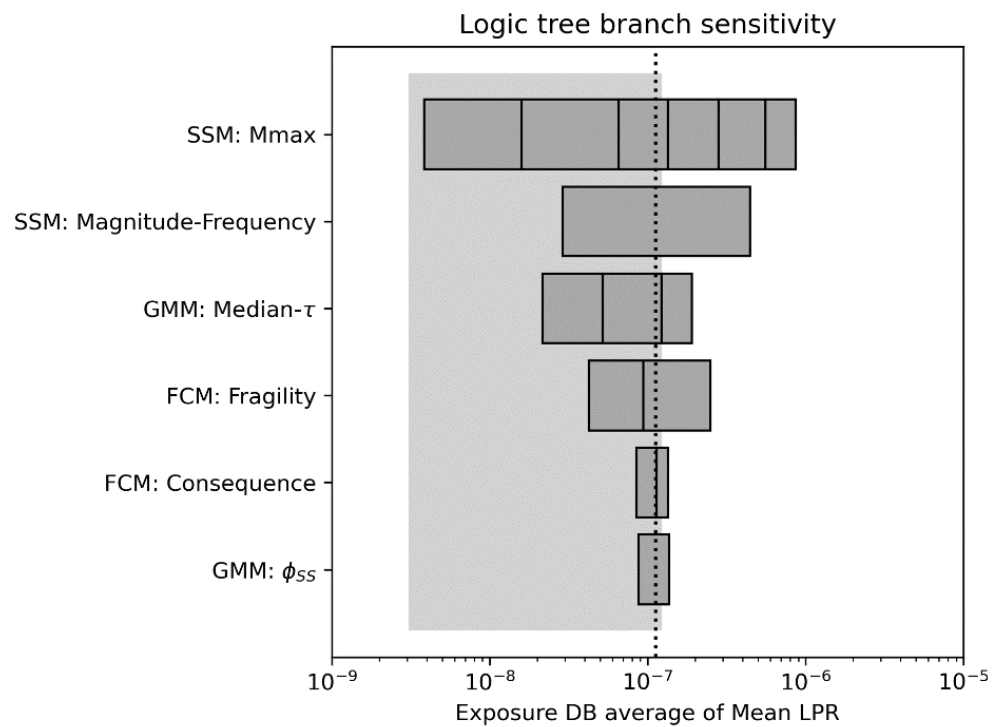


Figure 2: Sensitivity of the risk for each logic tree branch. Each black vertical line comprising the grey boxes is the risk for one model choice/value for that specific model setting. The dashed black line is the average mean risk from the pSHRA2021 [TNO, 2021a] after applying the weight for each logic tree branch choice/value. The P10-P90 bandwidth resulting from the logic tree is shown as a light grey band.

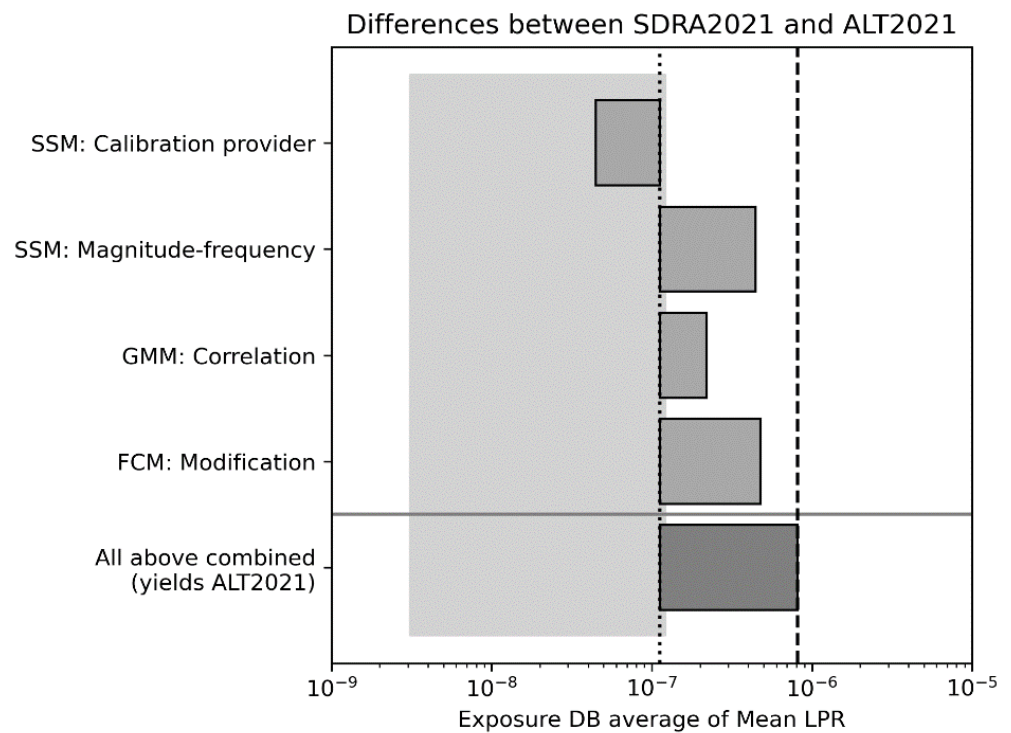


Figure 3: Sensitivity analysis results for three new model settings, conducted as part of the pSHRA2021 report, compared to the average mean risk from the pSHRA2021 (black dotted line) and P10-P90 bandwidth (light grey band). The newly tested model settings are the SSM calibration provider, GMM correlation, and FCM parameter variations. The change in settings is applied relative to the pSHRA2021, and a horizontal bar indicates the effect on the average LPR-value. For example, if the pSHRA2021 would be run using the period-to-period correlation structure of the sensitivity test, the average LPR-value of the pSHRA2021 would be higher ("GMM: Correlation").

2 Approach

We perform sensitivity analyses for a large set of model settings by varying one setting (parameter value/ input data/ implementation/ assumption) per test with respect to a set of default settings that constitutes the 'base case'. The base case is a variant on the pSHRA2021, with the same settings but a different implementation of the calibration module in the seismic source model (SSM). This calibration implementation is described in TNO, 2021b. The calibration used for the pSHRA2021 was supplied by NAM [TNO, 2021a] and cannot be reproduced nor validated by TNO on the basis of the documentation provided by NAM [TNO, 2021b]. Since every change in the SSM (parameters value/ input data/ implementation/ assumption) requires a re-calibration of the SSM, this renders the pSHRA2021 unusable as a base case; it would not be possible to uniquely attribute changes in results to changes in the sensitivity setting. The difference in the mean risk between base case and pSHRA2021 is approximately a factor 3, with the new base case's mean risk equal to that of the SSM calibration provider sensitivity test (Left side of the top horizontal bar in Figure 3).

Below, the tested model settings, their base case values/ implementation, and their tested values/ implementation are summarised per model chain component.

2.1 Risk metrics

The sensitivities of parameter/model choices on the risk relative to the risk of a base case will be quantified using the following metrics: The mean risk and the percentiles P10 and P90. We understand mean risk here as either i) the average risk over the entire building database (mean LPR of each building, averaged over the total building stock), or ii) the average risk over the 1% most at risk buildings (1594 buildings, Figure 4). These two mean risk measures provide insights on how the parameter/model choice sensitivities vary across the entire building distribution.

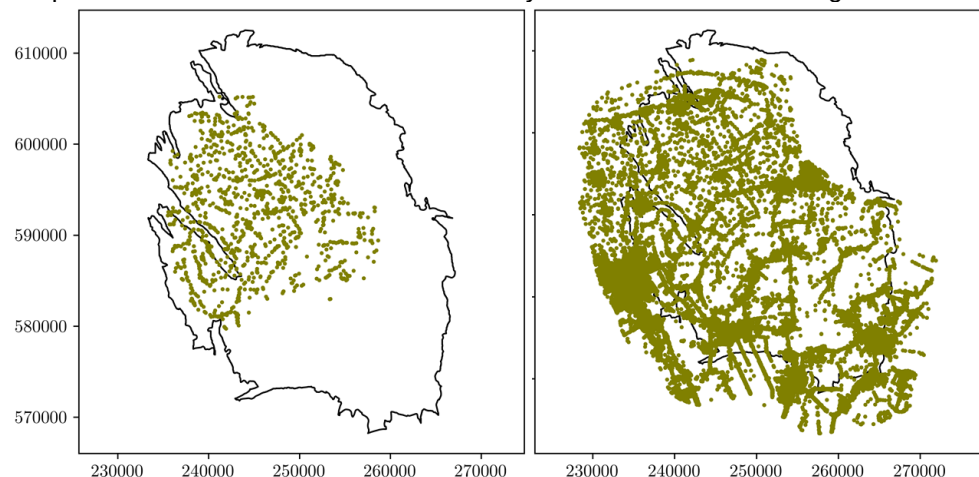


Figure 4: Left: Locations of the top 1% most at risk buildings on the contour of the Groningen gas field. An alternative risk metric is defined as the mean of these most at risk buildings. Right: Locations of all buildings in database.

The P10 and P90 percentiles are based on the epistemic uncertainties accounted for in the logic tree. The tested period of interest is the gas year 2021/2022, running from the 1st of October 2021 until the 1st of October 2022.

2.2 Seismic source model (SSM)

The most recent version of the seismic source model was used [TNO, 2020c]. The SSM is based on a covariate field representing what is known as the ‘smoothed incremental Coulomb stress field’ [Bourne & Oates 2017]. We varied (1) the deterministic input data from which the Coulomb stress field is computed, and (2) the implementation choices, i.e., how the covariate field is calculated. The covariate Coulomb stress field, together with prior probability distributions for a number of parameters, and an earthquake catalogue, is utilised to calibrate the seismicity rate predicted by the SSM. We explore the predicted seismicity (3) with alternative priors and variations in the earthquake catalogue.

The test cases are:

1. Uncertainty in deterministic input
 - a. *Reservoir thickness* (default = mean reservoir thickness of the geological model). The base case mean reservoir thickness is interpolated from well-derived thickness data and adjusted to the geometric model (based on PrSDM seismic interpretation). The observed residuals between modelled thickness and well derived thickness data can be best represented by a normal distribution with mean of 1.3 m with standard deviation of 5.4 m. The alternative reservoir thickness tested here is based on this residual distribution. To do so, a residual grid was computed by means of a Sequential Gaussian simulation with a Kriging interpolation range of 1.5 km (Figure 5, left panel). This residual grid is an estimate of possible local anomalies of the sedimentary distribution and is added to the mean reservoir thickness.
 - b. *Fault locations* (default = fault model NAM). The test case consisted of an alternative grid of fault locations. This alternative fault model is based on the original fault model supplied by NAM, supplemented by small offset faults based on Kortekaas & Jaarsma [2017] (Figure 5, right panel). These additional faults have been assigned a uniform offset of 8 m, approximately the minimum offset required to be detectable through seismic techniques.
 - c. *Pressure depletion grid* (default = pressure grid provided by NAM): Two alternative pressure grids are tested: 1) A base case grid augmented with a residual grid to represent the pressure uncertainty increasing with distance for the production locations. At production locations pressure residuals were assumed normal distributed (mean: 0 bar, standard deviation: 10 bar). Based on this input a residual grid was computed by means of a Sequential Gaussian simulation with a Kriging interpolation range of 10 km (Figure 6, center). 2) A grid mimicking the history match results of the V5 dynamic reservoir model (Figure 6, right). The current V6 dynamic reservoir model (introduced in 2018) substantially changed the history match boundary conditions. This is mainly reflected in the area’s with limited well (pressure) control. Yearly V5-V6 pressure difference maps were generated till 2017. Pressure differences from 2018 onwards were assumed equal to 2017.

- d. *Compaction coefficient grid* (base case = coefficient grid provided by NAM): We use an alternative compaction coefficient grid based on Groningen field porosity (Figure 7). This grid has recently been calibrated using subsidence levelling data [NAM, 2021].

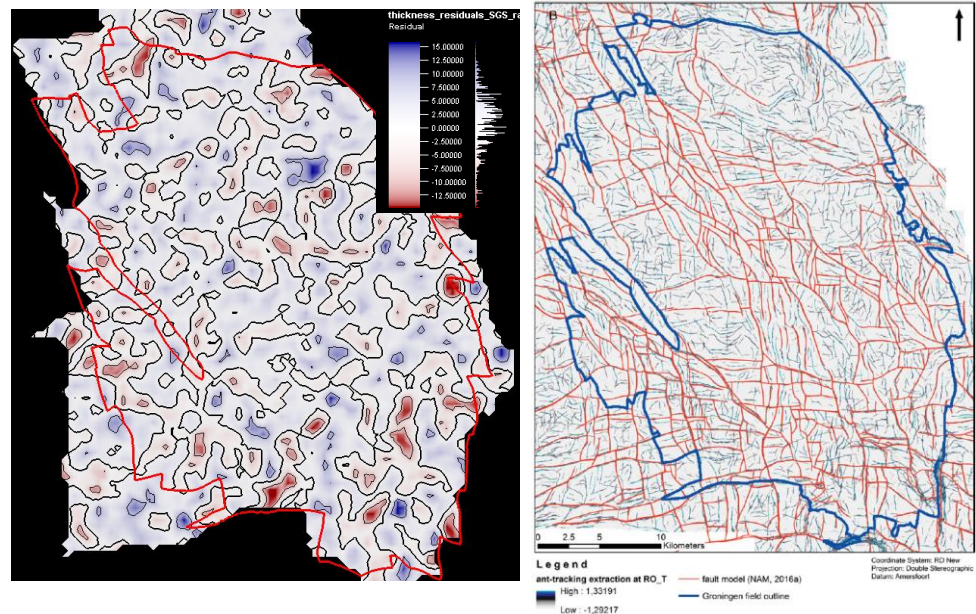


Figure 5: Alternative deterministic input. Left: Residual thickness grid (approximate Groningen field outline in red) that is added to the mean thickness grid (base case) to produce the alternative reservoir thickness grid. Right: Seismic attribute map of the top of the Rotliegend formation (approximate Groningen field outline in blue) from Kortekaas & Jaarsma [2017]. Based on this map, additional faults (blue segments) were identified and added to the base case NAM fault model (red fault segments).

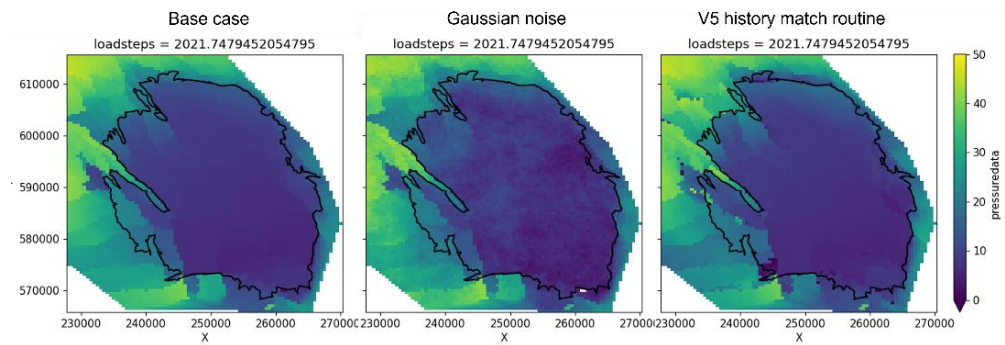


Figure 6: Pressure grids for the 2021 gas years: Base case (left), base case augmented by residual noise (center), and mimicking V5 history match results. Units are in bar. Groningen field contour in black.

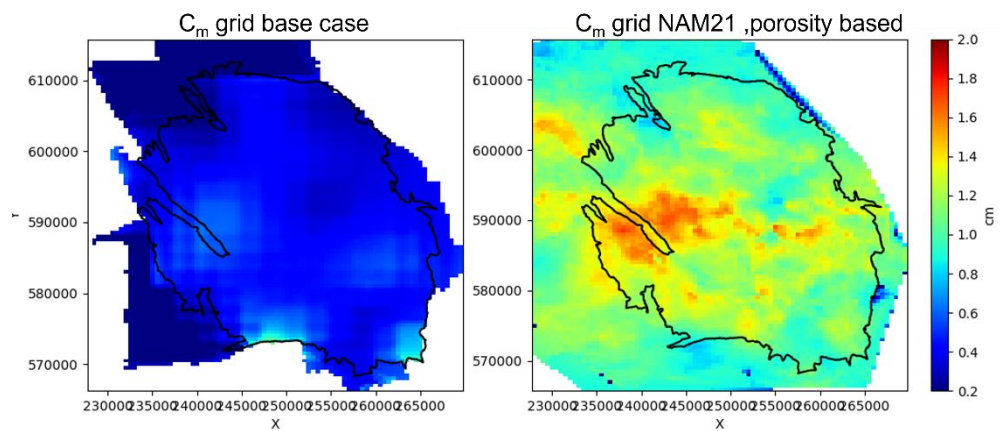


Figure 7: Compaction coefficient grids: Base case grid (left) and a porosity-based calibrated grid (NAM 2021). Units are 10^{-5} Pa^{-1} .

2. Model implementation on Dynamic Subsurface Model (DSM) Coulomb stress: The computation of the incremental Coulomb stress field (ΔC) requires a number of steps to be taken. The 'TNO' implementation is based on the scientific publications by NAM [Bourne & Oates 2017; Bourne *et al.*, 2018], but without access to the NAM code base itself. After this implementation was completed, TNO was asked to reproduce an 'exact' copy of the NAM model, for which access to the NAM code base was arranged. The 'NAM' implementation that resulted from this indeed reproduces the ΔC -field produced by the NAM code base to an arbitrary number of machine precision. To get to this exact replica, many implementation changes had to be made with respect to the 'TNO' implementation. These changes include, but are not limited to:

- Methodology of extrapolation to field locations without data (e.g. thickness, pressure, compressibility).
- A pre-smoothing step of the pressure depletion field with a fixed smoothing kernel length of 5000 m, independent of the model parameter σ , which is used to smooth the ΔC -field.
- Implementation of a custom Gaussian smoothing kernel with non-standard boundary conditions.
- An alternative method for mapping the unstructured fault data to the ΔC spatial grid.

- Removal of the $\rho(x)$ term in the calculation of ΔC .
- Use of a different value for γ (this only results in a uniform scaling of ΔC).
- The use of the full $h(x)$ field, rather than reservoir thickness at fault.

As a result, the TNO Model Chain code base includes a 'TNO' implementation and a 'NAM' implementation for the calculation of ΔC . In addition, the 'NAM' implementation computes a single ΔC field for the calculation of activity rate, and an independent single ΔC field for the magnitude model. The TNO implementation uses consistent ΔC fields for both the activity rate and magnitude model, and integrates over a range of ΔC fields to account for uncertainty in the covariate.

The base case uses the TNO implementation and integration over ΔC . We study the impact of these different choices in the following sensitivity runs:

- No integration of the Coulomb stress.* The test case is not to integrate the Coulomb stress field, while using the TNO implementation of the ΔC calculation.
 - NAM implementation of ΔC calculation.* The test case uses the NAM implementation of the ΔC calculation, rather than the TNO implementation. Since the choice of prior is ambiguous for this case, we tested two alternative priors.
 - Coulomb stress and seismicity rates are calculated on the faults.* In the base case, the seismicity rate is calculated from the spatially smoothed incremental Coulomb stress field ΔC . As a result, the seismicity rate is not calculated on a population of faults, but on a spatially smooth field. The theoretical background of the source model is based on seismicity occurring on faults. To test the sensitivity to this choice, we calculate ΔC on the faults, and subsequently calculate seismicity on the faults. This seismicity field is then spatially smoothed before comparison to the data. A natural consequence of this change in the source model is that the magnitude-frequency model and the activity rate model no longer have independent parameter spaces. This implementation was therefore only combined with a constant b-value model. Therefore, we will compare this case not with the base case, but with the test case with a constant b-value model (see below).
- Magnitude-frequency (MF) model
Two magnitude-frequency models are used in the base case: The hyperbolic tangent model and the taper model. These two models constitute the MF-branch of the logic tree (Figure 1), and utilize the Coulomb stress field as the covariate. Here, we test three other magnitude-frequency models:
 - Constant b-value model: Here, the MF is simplified to the exponent b as the single parameter. Parameter b is independent of a predictor (covariate), and is therefore constant across space and time. A range of b-values is provided in the prior.
 - Linear thickness model: Exponent b is linearly dependent on the reservoir thickness (provided as standard deterministic input) by a slope parameter. Both b-values and slope are provided as a range in the prior. Recent work by TNO showed that, together with the thickness-dependent step model (see below), the thickness-dependent

linear model is the best predictor for magnitude-frequency distributions [Kraaijpoel et al., *under review*].

- c. Step thickness model: Two b-values are used to construct an MF: One for thinner parts of the reservoir, one for the thicker parts. The transition reservoir thickness from one b-value to the other (the 'step') is the third parameter defined in the prior.

Note that the alternative MFs are truncated at the M_{max} values defined in the logic tree branch.

4. Seismic source model calibration

- a. *Number of free parameters in the prior* (default = 3 free parameters). There are three free parameters that link the covariate field to the observed seismic activity rate. One of these parameters is material parameter H_s , which is defined as a function of the bulk modulus of the solid skeleton material and Poisson's ratio. In the model, H_s primarily determines the extent to which Coulomb stress changes correlate with reservoir strains. In the test case, H_s is set to a constant value of $5.0 \log_{10} (\text{MPa}^{-1})$. This value corresponds to the most likely value for H_s in the base case.
- b. *Calibration date range* (default = 1995-2020). All earthquakes falling in this time interval are used to calibrate the model. Alternative date ranges were tested:
 - i. 2000-2020
 - ii. 1995-2019
 - iii. 1995-2018
 - iv. 1995-2021/11/16.

The date range for the last case includes the larger magnitude events observed in October and mid-November 2021. This simulation produces a partial 'hindcast' on the period 2021.
- c. *Earthquake aftershocks using ETAS* (default = yes). In the base case, aftershocks resulting from a main earthquake event (i.e., those used for calibration of the model) are predicted using the Epidemic-Type Aftershock (ETAS) model [Ogata 1988], and thus contribute to the forecasted seismicity. In the implementation, the first event is assumed to be the main event. All events (main & background) are assumed to have (and are drawn from) the same distribution. In the test case, the ETAS model is not used, and the main/background event rate produces all the forecasted seismicity.
- d. *Range of included earthquake magnitudes* (default: $M_{min} = 1.5$ [Dost et al., 2012]). Catalogued earthquakes below the magnitude threshold M_{min} are not used for calibration of the SSM. According to Varty et al. [2021], the completeness of the Groningen earthquake catalogue is near $M_{min} = 1.2$. We therefore tested M_{min} values ranging between 1.2 and 2.0, with steps of 0.1 (9 tests, including base case).
- e. *Alternative earthquake catalogue* (default = KNMI catalogue induced_earthquakes). The test case uses an alternative catalogue obtained from the KNMI website (<https://datapatform.knmi.nl/dataset/aardbevingen-catalogus-1>). This catalogue differs from the base case catalogue by magnitudes that are rounded to two decimals instead of one, and the earthquake source locations are slightly different (Figure 8).
- f. *Consistent event filtering contour*
The spatial contour used to filter which earthquake source locations are included in the SSM calibration is in this test case used for the event rate

prediction as well. In the base case, the predicted events could lie outside of this spatial contour.

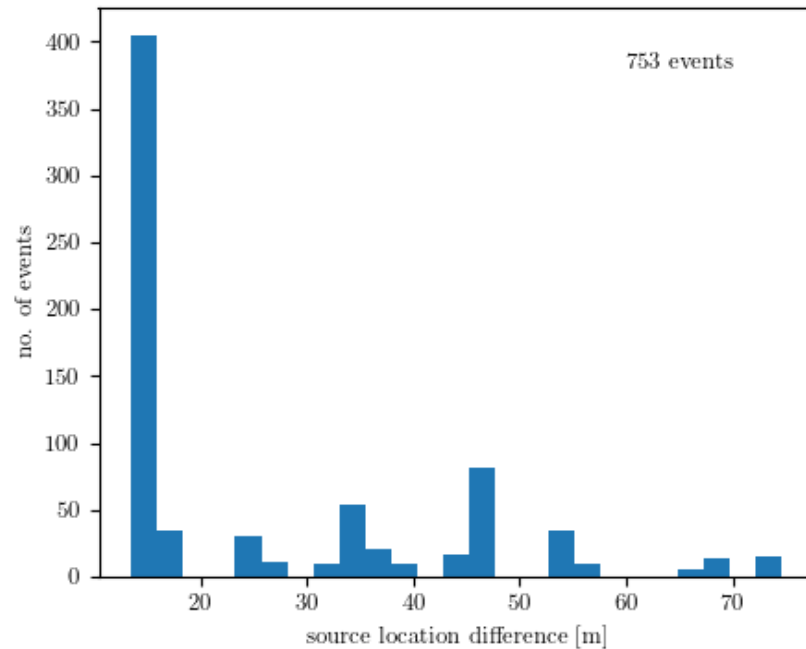


Figure 8: Histogram of the difference in source locations in catalogues induced_all.csv and catalogue_KNMI_contour.csv, for the same earthquake events.

2.3 Ground motion model (GMM)

There has been some discussion on the period-period correlation model of the site-to-site variability in the site response amplification model [TNO, 2020b; 2021b]. For the base case the prescription of the pSHRA was used. This prescription assumes that there is no period-period-correlation in the site response. With respect to the base case we discern three alternative choices: The first was already applied and reported in the sensitivity analysis of the pSHRA 2021 [TNO, 2021a] (Figure 3): it assumes that the period-period correlation model for the reference level ground motion is applied consistently in the site response model. Alternatively, two variations are included with perfect (full) period-period correlation in the site response, as this appears to be the prescription in the forthcoming GMM v7. The full correlation is tested both with a complete log-normal distribution of the site-to-site variability, and with a 3-point approximation according to the GMM v7 documentation [Bommer et al., 2021].

2.4 Fragility and consequence model (FCM)

Four FCM sensitivity test runs were conducted. The FCM sensitivity test runs reflect the model choices TNO proposed in the SHRA status report 2022 [TNO, 2021b]. The combined sensitivity of these four test runs was already published in the sensitivity chapter of the pSHRA 2021 report [TNO, 2021a]. The sensitivity tests include alternative parameters for the unreinforced masonry (URM) typologies only (Table 1):

1. URM model uncertainty (fragility): increased model uncertainty (β_m) from 0.27 (default value) to 0.35. Here we follow the FEMA model uncertainty classification of superior models (MDOF; multi degree of freedom) versus average models (SDOF; Single degree of freedom). The assigned increased model uncertainty accounts for the observed (TNO, 2020e) difference in variability between SDOF and MDOF derived fragility curves.
2. Fragility parameters for URM classes: The model based fragility curve is considered as 'best estimate'; it is consequently implemented as the middle branch of the logic tree (see Figure 9). In the base case the most fragile typologies were implemented as the 'upper branch', because the model development team [Crowley et al., 2019] believed that the models underestimated the true building strength.
3. Consequence parameters for URM classes: (Comparable approach as fragility) The model based consequence curve is considered as 'best estimate'; it is consequently implemented as the middle branch of the logic tree. In the base case the most fragile typologies were implemented as the 'upper branch'.
4. Shift median seismic capacity (fragility): A -15% shift in seismic capacity was applied for all URM types, whereas in the base case this is only applied to the most fragile typologies. A variation study of between building and within building variations (TNO, 2020e) showed this shift in seismic capacity. This is expected because the fragility tends toward the weakest point within the variations.

Table 1: Detailed settings for the four FCM sensitivity runs, specified per URM vulnerability class

vulnerability class	Model uncertainty β_q		Representative logic tree branch fragility curve		Representative logic tree branch consequence curve		Shift median seismic capacity	
	Base case	FCM-1	Base case	FCM-2	Base case	FCM-3	Base case	FCM-4
URM1F_B	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM1F_HA	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM1F_HC	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM2L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM3L	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM3M_B	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM3M_D	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM3M_U	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM4L	0.1	0.25	upper	middle	upper	middle	-15%	-15%
URM5L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM6L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM7L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM8L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM9L	0.1	0.25	middle	middle	middle	middle	0%	-15%
URM10	0.1	0.25	middle	middle	middle	middle	0%	-15%

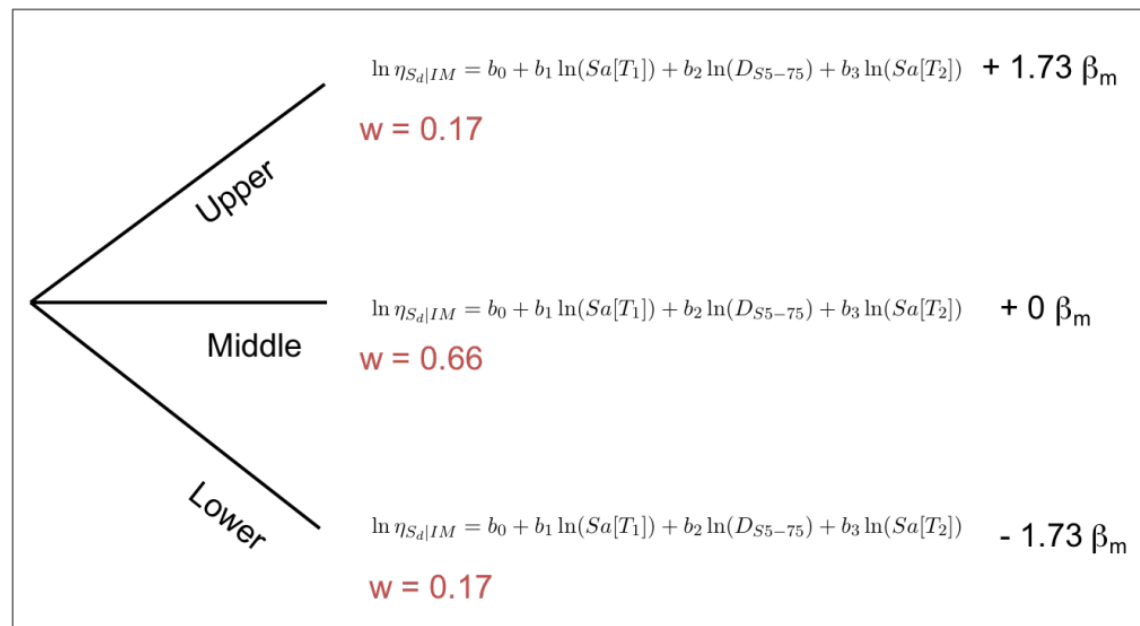


Figure 9: Logic tree for the fragility model [Crowley *et al.*, 2019] to incorporate the model uncertainty β_m .

2.5 Exposure database (ED)

The buildings in the base case exposure database have a probability for each typology that they may be assigned to. In the alternative exposure database tested here, each building has been assigned fully to its most probable typology.

3 Results

3.1 Overview

The absolute mean risk values and the P10-P90 range are shown in Figure 10. For a convenient comparison between tests we present the mean risk values relative to the base case in Figure 11. The mean risk exceeds the P90 percentile value for the majority of test cases, which shall be discussed more thoroughly in section 7.

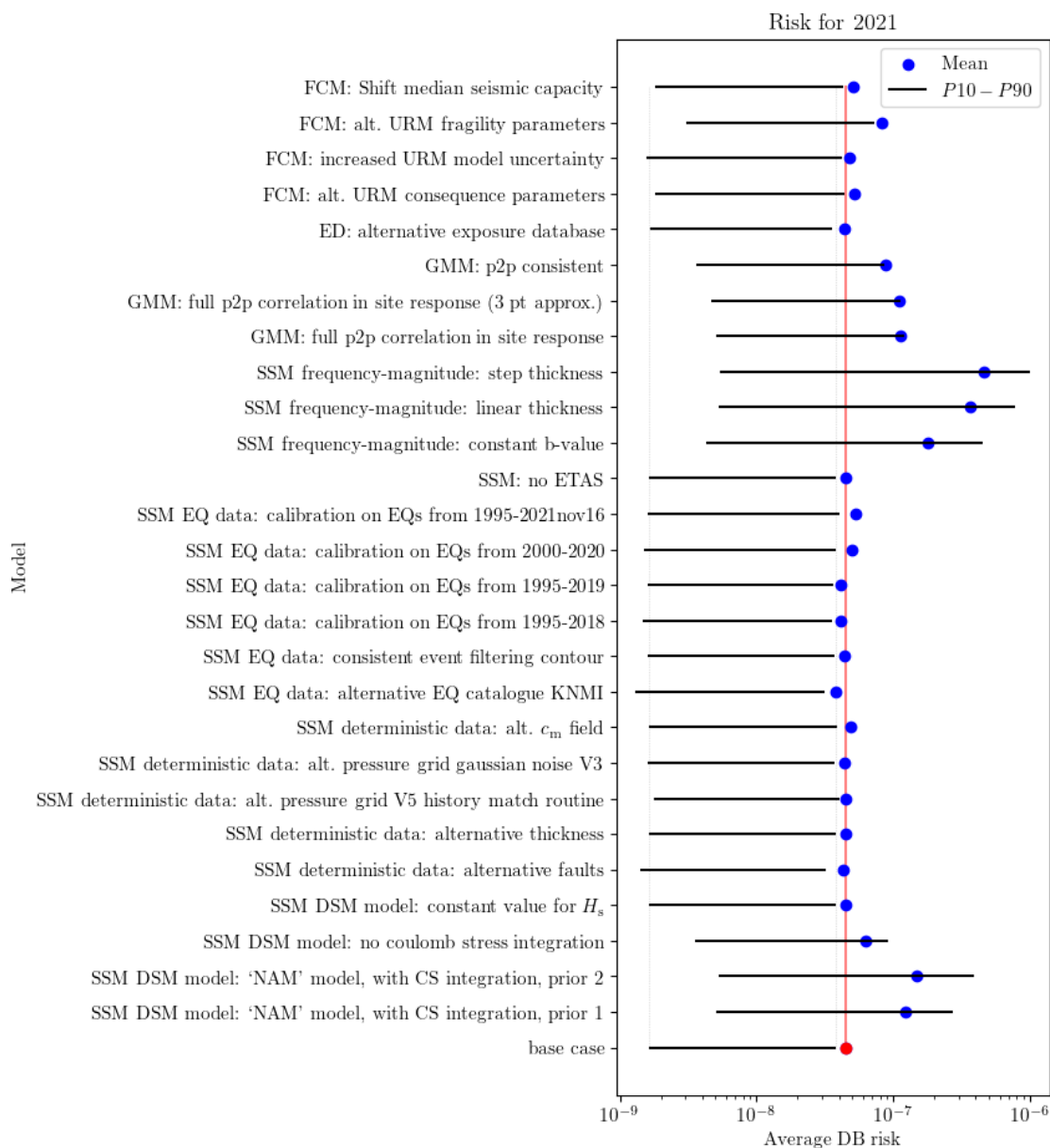


Figure 10: Simulated mean risk (blue markers) and P10-P90 range (black bars) for the tested cases. Mean risk of the base case indicated by the red marker and line, and reference of the P10-P90 of the base case highlighted by the dashed grey lines.

SSM DSM variations: Variations in the dynamic subsurface model (DSM) in the SSM module show that the largest changes in mean risk occur with variations for the model implementation for computing the covariate Coulomb stress field (the test with *no Coulomb integration* and the two “NAM” style covariate models, Figure 10). The mean risk is between 1.42 (no Coulomb integration) to 3.35 (‘NAM’ with prior 2) times higher than the base case. Note also that for these higher risk tests, the mean falls between the P10 and P90 values. This indicates that the shape of the risk probability distribution is less skewed for these cases relative to the base case. The DSM implementation for Coulomb stress on faults is discussed in section 3.4. Last, a constant value for prior parameter H_s results in a negligible change in mean risk.

SSM earthquake calibration: Changes in the catalogue of earthquakes used to calibrate the SSM have an impact on the risk by a factor 0.85 up to 1.45: First, rounding the earthquake magnitudes from one decimal (base case) to two decimals (the KNMI catalogue), and a small change in source locations, results in a risk that is 0.85 times lower than the base case. This change is a combination of three factors: i) a small change in source locations, ii) a changed catalogue of magnitudes to calibrate on, and iii) an effective change of included earthquake magnitudes. This third point needs some explanation: For the base case (magnitudes rounded to one decimal, $M_{\min} = 1.5$), events with an unrounded magnitude 1.45... and higher are included in the calibration, as they are rounded to 1.5. For the same $M_{\min}$, the smallest unrounded magnitude events included in the alternative case are 1.495..., as these are rounded to 1.50. Note that the rounded magnitudes are already provided as input, and the unrounded magnitude is unknown. Thus, it is likely that fewer earthquakes are included for calibration in the test case relative to the base case. This could be interpreted as a tightening of the “effective” $M_{\min}$ closer towards M1.5. The effect of rounding has been accounted for in the SSM through a different bin size. Second, changing the calibration time interval may result in a lower risk (0.93 times the base case) when disregarding the last year or two years of the catalogue. Disregarding the first 5 years of induced earthquakes (leaving a catalogue from 2000-2020) increases mean risk by a factor 1.13. Including earthquakes up to the 16th of November 2021 increases risks by a factor 1.20. Further details on the date ranges are given in section 3.2. Varying the range of included earthquakes prior to calibration by excluding events with magnitudes below different values for $M_{\min}$ increases the risk up to a factor 1.45 (Figure 13). This is further detailed in section 3.3. The oversight of an inconsistent spatial filtering contour for predicted events (i.e., the base case) reveals very little difference with respect to the test case with consistent application of the contour.

SSM deterministic data: The mean risk changes resulting from variations in deterministic data are very minor for alternative thickness, alternative faults, and alternative pressure grids (around a factor 0.99, Figure 11). The effect of an alternative compaction coefficient grid increases the risk by a factor 1.05.

SSM ETAS: Switching ETAS off in the test case results in a vanishingly small change in mean risk compared to the other test cases.

SSM Magnitude-Frequency model: The choice of MF causes the largest changes in mean risk with respect to the base case, for all tested cases. A constant b-value

model increases the risk by a factor 4, whereas the linear- or step thickness MFs increase the risk by a factor 8.2 and 10.2, respectively. Further details on the MF distributions are given in section 3.4.

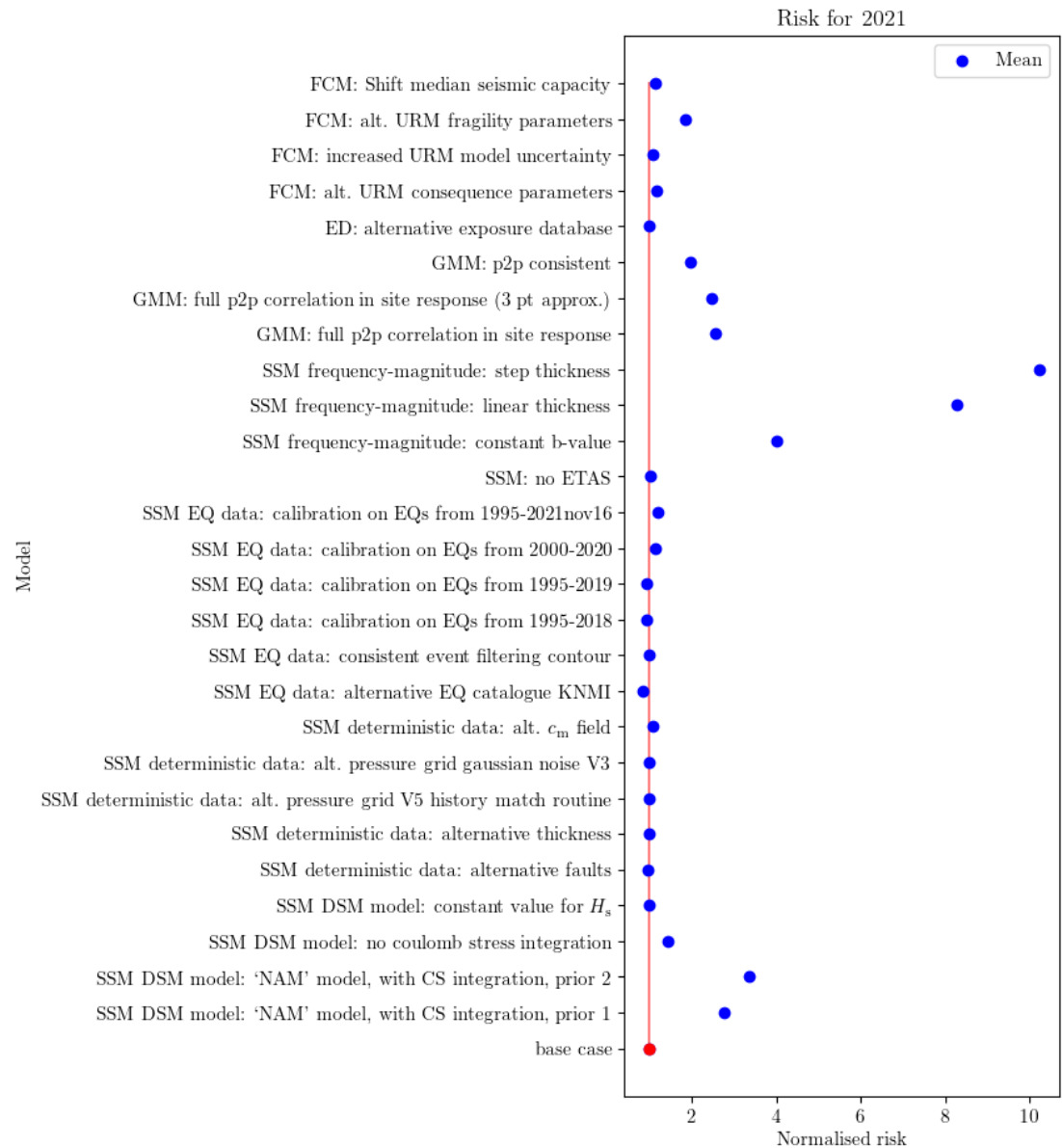


Figure 11: Simulated mean risk normalised by the base case mean risk (red marker and line) for the test cases described in section 2.

GMM: The tests in the GMM module yield higher mean risks and reshape the probability distribution (illustrated by P10-P90 relative to the mean risk) (Figure 10). Using a full p2p correlation in site response (either using a 3-point approximation or not) increases the mean risk by around a factor 2.4. A consistent p2p increases risk by a factor 1.88.

Alternative exposure database: The choice of an alternative exposure database (ED) hardly impacts the mean risk.

FCM: Finally, changes made in the fragility and consequence model (FCM) result in increased mean risk: A median shift and an alternative model uncertainty increase the risk by a factor 1.07 and 1.14, respectively. Alternative consequence parameters increase the risk by 1.18, and alternative fragility parameters increase the risk by a factor 1.83.

3.2 SSM: Calibration date range

The changes in mean risk with calibration date range arise in the calibration module of the SSM. Here, the earthquake magnitude and/or the earthquake frequency may vary from the base case. Figure 12 shows that, when recent earthquakes are not used for calibration (i.e., 1995-2018 and 1995-2019), the forecasted number of events in gas year 2021/2022 is lower for all magnitudes. When the oldest 5 years are excluded (2000-2020), we predict more events with a magnitude above 2.2. Including the most recent 10.5 months of induced earthquakes in the calibration results in a higher frequency of earthquakes for all magnitudes. It should be noted that the latter result is biased in the sense that the very experiment was inspired by the recent occurrence of a relatively high number of events. Closing the catalogue just after a peak of activity is expected to result in an increased risk (*ceteris paribus*).

The (relative) frequency-magnitude curves in Figure 12 follow a largely linear trend for most of the magnitude range. This trend is caused by a change in *b*-value. The valleys and hills that appear in the range between magnitudes 3 and 5 are caused by a change in the taper position. A change of the taper to lower magnitudes results in a valley, a change to higher magnitudes results in a hill.

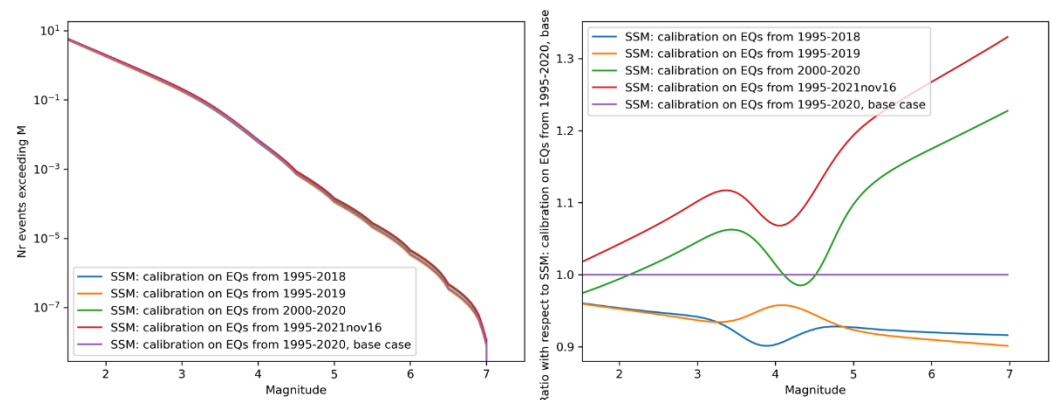


Figure 12: Left hand panel: Magnitude-frequency distributions for the tested calibration date ranges. Right hand panel: MF models normalised by the base case (calibration range 1995-2020).

3.3 SSM: Calibration range of included magnitudes

The mean risk as a function of included magnitudes over the range of $M_{\min} = 1.2$ to $M_{\min} = 2.0$ follows a remarkable trend (Figure 13). In the tested range, we can distinguish two minima: One absolute minimum at $M_{\min} = 1.5$ (coincidentally the base case value) and one local minimum at $M_{\min} = 1.8$, with a local maximum at $M_{\min} = 1.7$. This risk at this local maximum is 1.35 times higher than the base case.

At the extreme ends of the tested range, we observe mean risks that are 1.45 and 1.40 times higher for $M_{\min} = 1.2$ and $M_{\min} = 2.0$, respectively.

The P10 values (left side of the black bars in Figure 13) increase with increasing values for $M_{\min}$, and do not reveal local minima or maxima. In contrast, one absolute minimum is observed in the P90 values, located at the base case.

As noted before, the alternative earthquake catalogue from KNMI with 2-decimal magnitudes instead of 1-decimal magnitudes can be interpreted to change the effective $M_{\min}$ as well, even though we maintain $M_{\min} = 1.5$. In the base case, earthquakes classified as M1.5 may have unrounded magnitudes of 1.45 to 1.55. For the 2-decimal test case, earthquakes classified as M1.50 may have unrounded magnitudes of 1.495 to 1.505. Hence, the rounding difference results in a 'tighter' value of $M_{\min}$. The lower risk (0.85 times the base case) cannot be described solely to the tighter $M_{\min}=1.5$ due to rounding, but also to different magnitudes across the entire magnitude range, and slightly different locations in the alternative catalogue. Nonetheless, there is some suggestion that the 'tightened' completeness near M1.5 is nearer to the risk minimum in the $M_{\min}$ value range.

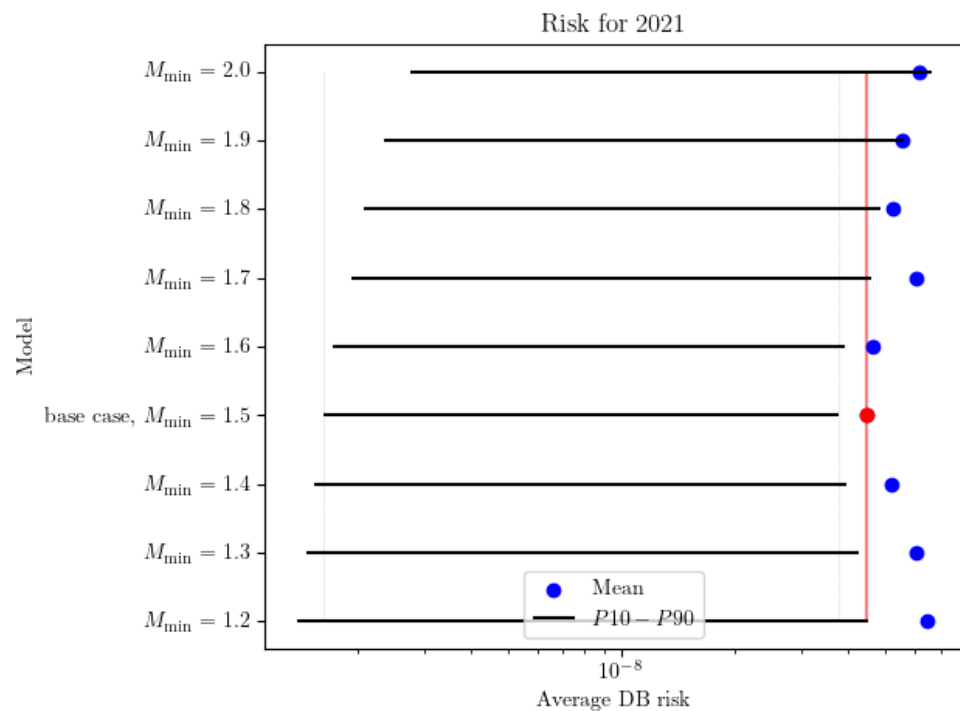


Figure 13: Mean risk and P10-P90 values for a range of $M_{\min}$ values. Base case ($M_{\min} = 1.5$) indicated in red.

With the exception of $M_{\min} = 1.6$, the predicted frequency-magnitude distributions show an increased frequency for events with magnitude greater than M2.5 (Figure 14). The linear portions of the relative frequency-magnitude curves all indicate a decrease in b-value (more high-magnitude events relative to low-magnitude events) relative to the base case. The valleys and hills, indicative for the taper position, suggest that for $M_{\min} < 1.5$, the taper shifts to lower magnitudes, and for $M_{\min} > 1.5$, the taper shifts to higher magnitudes.

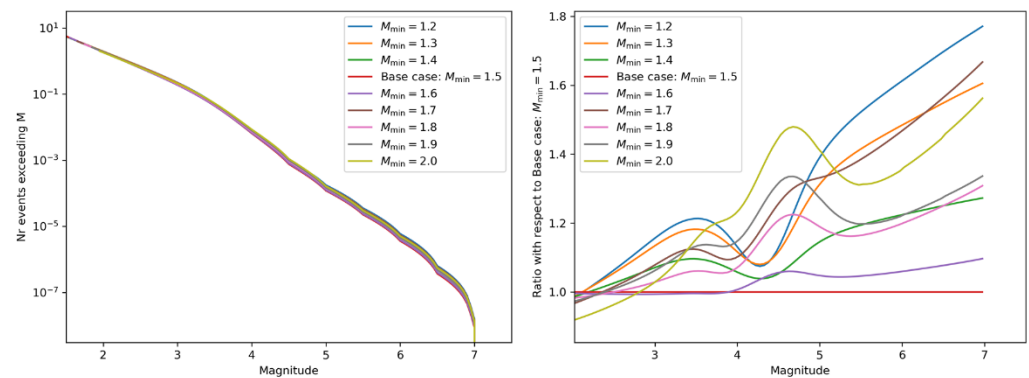


Figure 14: Left hand panel: Magnitude-frequency distributions for the $M_{\min}$ test cases. Right hand panel: MF models normalised by the base case ($M_{\min} = 1.5$).

3.4 SSM: Magnitude-Frequency models

The risk increases for all alternative MF models with respect to the base case model. This is a direct effect of an increased frequency for magnitude events exceeding M3.5 (Figure 15). For comparison, the MF distributions for the taper- and hyperbolic tangent models, both part of the existing MF logic tree branch, are shown as well.

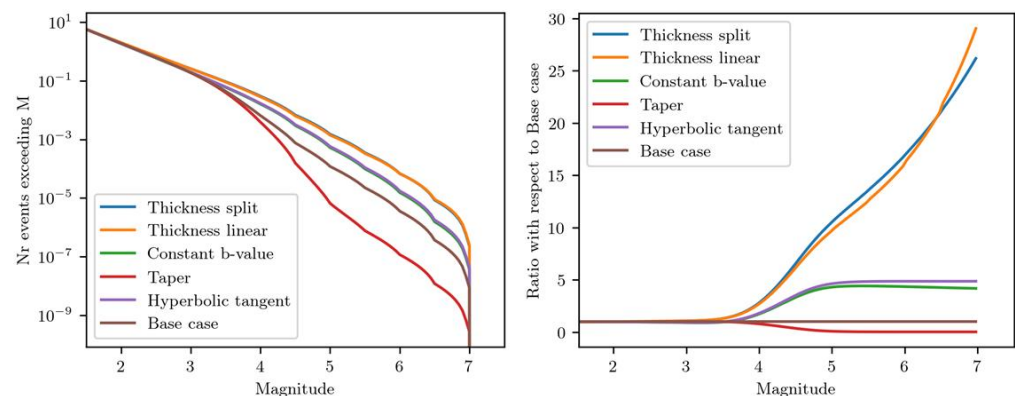


Figure 15: Left hand panel: Magnitude-frequency distributions for the base case (brown curve, 80% taper and 20% hyperbolic MF models), the Coulomb stress dependent taper (red curve) and hyperbolic tangent (lavender curve) models, the thickness dependent linear- and step models (blue and orange curves, respectively), and the constant b-value model. Right hand panel: MF models normalised by the base case.

3.5 SSM: On-fault Coulomb stress vs. smoothed Coulomb stress

The Coulomb stress calculated on the faults cannot be compared directly with the base case because in addition to a different Coulomb stress implementation, the magnitude-frequency model is different as well; a constant b-value model was used for the test case. We therefore compare the results of the on-fault Coulomb stress with the test case of a constant b-value model (Figure 16). The risk increases by a factor 1.34 relative to the case with a constant b-value model.

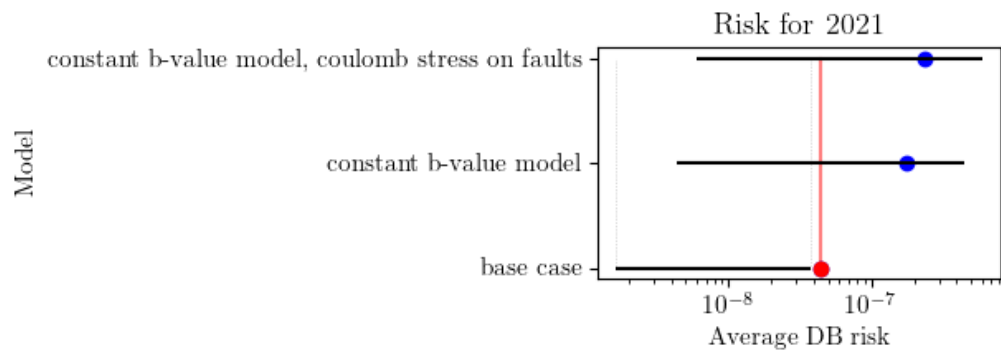


Figure 16: Risk for on-fault Coulomb stress and smoothed Coulomb stress, both cases with a constant b-value model. Base case (red marker and line) plotted for reference.

3.6 Alternative risk measure

The mean risks for the 1% most at risk buildings resulting from the base case settings and test settings are shown in Figure 17 (left hand panel). The absolute risk for all cases is an order of magnitude higher relative to the mean risk calculated over the entire building dataset (Figure 10), which is expected from including fragile buildings only. Note also that the mean risk for all test cases is now lower than the P90 value.

More informative is the potential shift in normalised risk change for the 1% risk metric with respect to the normalised risk change of the entire database (Figure 17, right hand panel). Although most normalised risks change very little, a few changes are highlighted:

- i. The normalised 1% risk metric for alternative URM fragility parameters is lower than that for the entire database. However, for an increased URM model uncertainty the normalised risk increases.
- ii. The increase in normalised risk from magnitude-frequency model alternatives is smaller for the 1% risk metric.
- iii. Some alternative deterministic data test cases yield different normalised risks: An alternative pressure grid decreases the normalised risk for the 1% most fragile buildings, whereas the risk increases for the alternative fault case.

On the whole, we may conclude that the relative variations in risk for the entire building database and the variations for the 1% risk metric are similar in magnitude, barring a few exceptions.

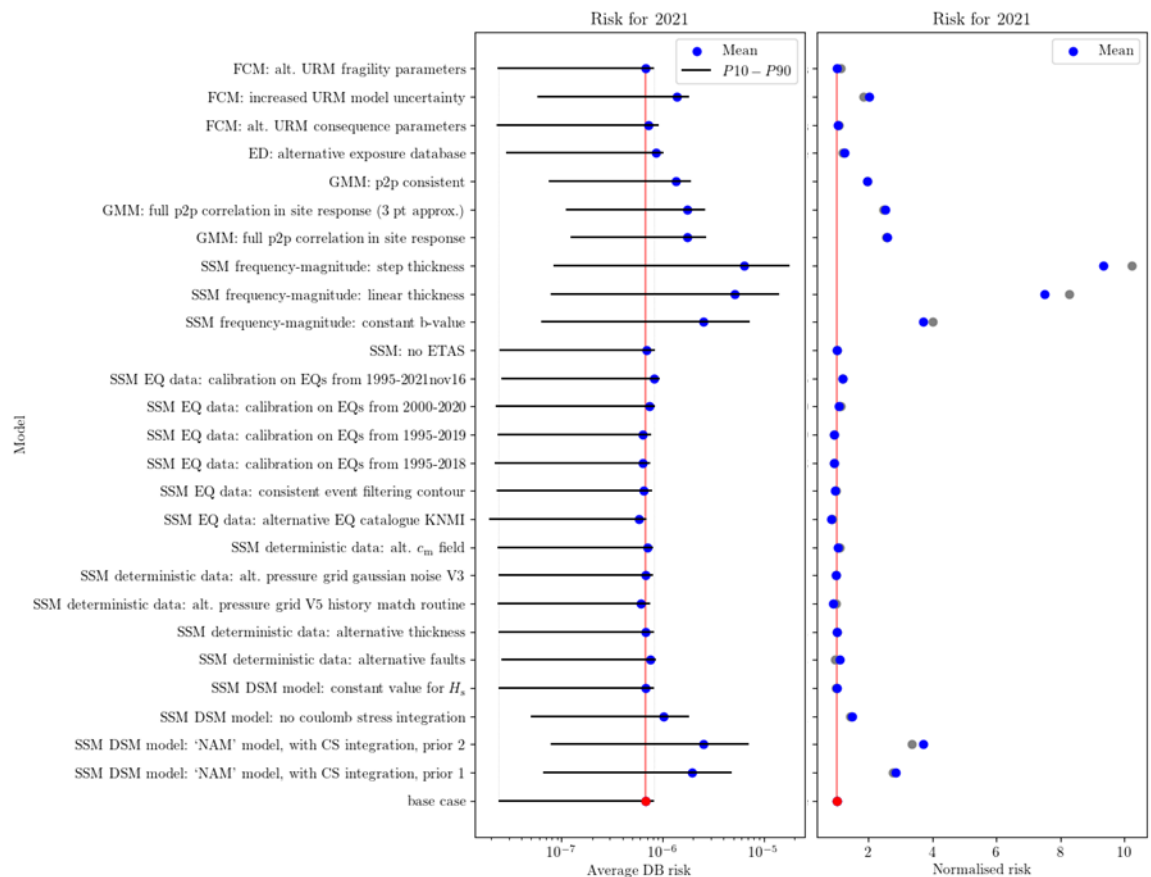


Figure 17: Absolute (left) and normalised (right) mean risk for the 1% most fragile buildings (blue markers). Grey dots on the right show the normalised mean risk over the entire building dataset.

3.7 Dissection of risk measure: mean risk, P10 and P90, and LPR of individual buildings

The majority of the tested cases, and the base case itself, yield a mean risk that falls outside the P10-P90 range (Figure 10), which at first glance may seem counter-intuitive. Here we provide some background on the probability distribution from which the mean and percentile values were taken (Figure 18). The probability distribution consists of all 1008 logic tree leaves (i.e., unique paths through the logic tree), weighted by the logic tree branch weights. This yields a weighted probability mass function (Figure 18, left side) with a high probability mass at low risk and a long tail towards higher risk consisting of lower probability mass leaves. The centre of mass, widely known as the mean, is found at right hand side of the highest mass probabilities (Figure 18, inset), where it balances the moment (mass times LPR-distance, in analogy with moment in physics) of the heavy bulk and the moment of the tail.

The percentiles P10 and P90 are located at a lower risk than the centre of mass. This is best illustrated in a cumulative mass plot (Figure 18, right side), which shows that 90% of the mass probability falls below the mean value. Note that, due to the variation in logic tree weights, the number of logic tree leaves per 10% percentile are different. For instance, 326 logic tree leaves fall in the P90-P100 range, whereas 119 leaves fall in the P0-P10 range.

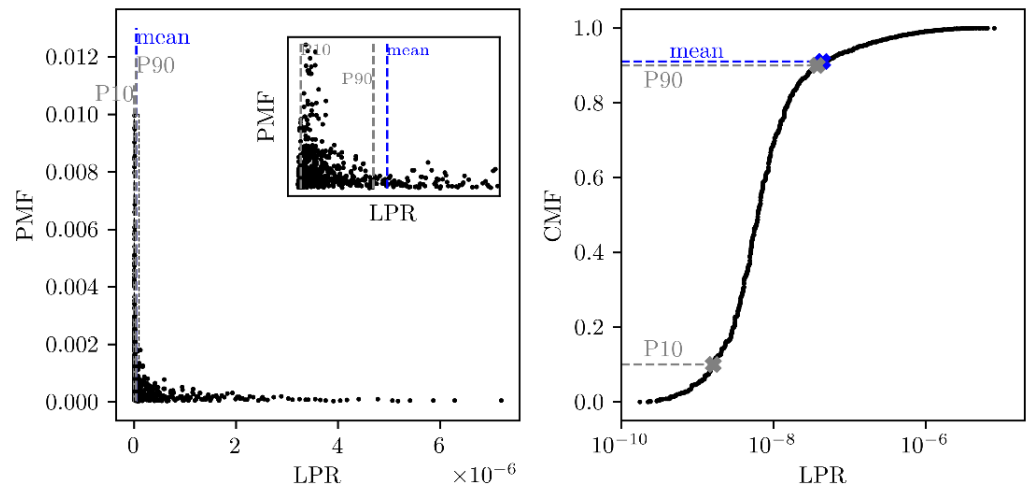


Figure 18: Weighted posterior mass function (PMF, left-hand panel) and weighted cumulative mass function (CMF, right-hand panel) versus LPR for each logic tree leaf (1008 in total) in the base case. Note the logarithmic horizontal scale on the right. The mean (blue) and P10 and P90 (both grey) risks are highlighted on the left and in the magnified inset. The percentiles are indicated on the right-hand side, showing that the mean falls in a percentile exceeding 0.9 (P90).

3.7.1 Highest & lowest risk logic tree leaves

The distribution over the logic tree branches of the 326 highest risk logic tree leaves that constitute the >P90 part of the risk uncertainty is visualised in Figure 19. The majority of higher risk leaves include the hyperbolic tangent magnitude frequency model (MF- β). Higher risk is also correlated to higher $M_{\max}$ values. There seems to be a slight preference for the upper branches of the median GMM and fragility branches. The high risk branches are distributed equally over the GMM variability and Consequences branches.

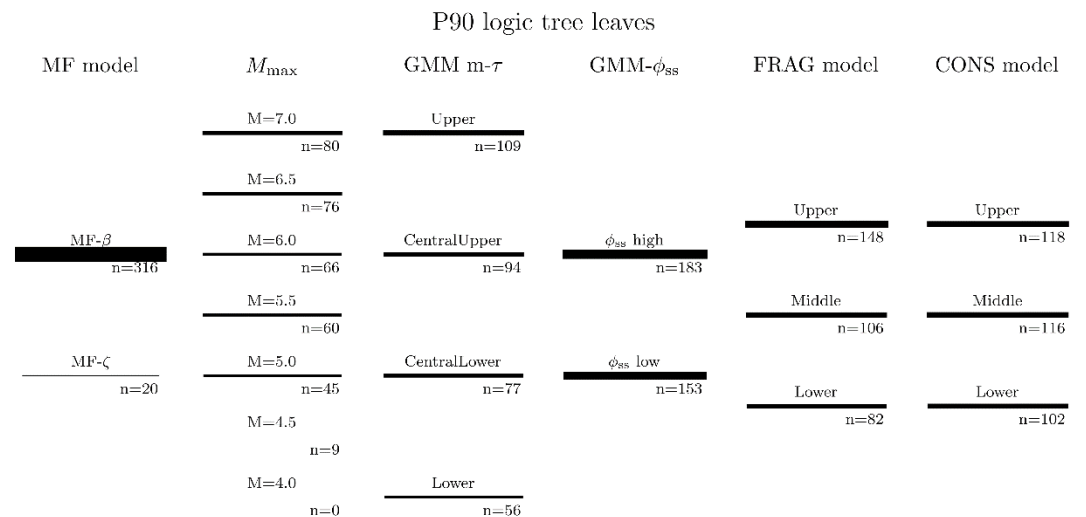


Figure 19: Logic tree showing how the >P90 leaves are distributed over the logic tree branches. Number of leaves on each branch given by n . Line thickness of each branch proportional to n . See Figure 1 for logic tree weights.

For the 119 lowest risk leaves (Figure 20), we can identify more or less the opposite trends for the $M_{\max}$, MF model, median GMM and Fragility branches. In addition, we see that the low risk leaves are more abundant in the lower branches of the GMM variability and consequence branches.

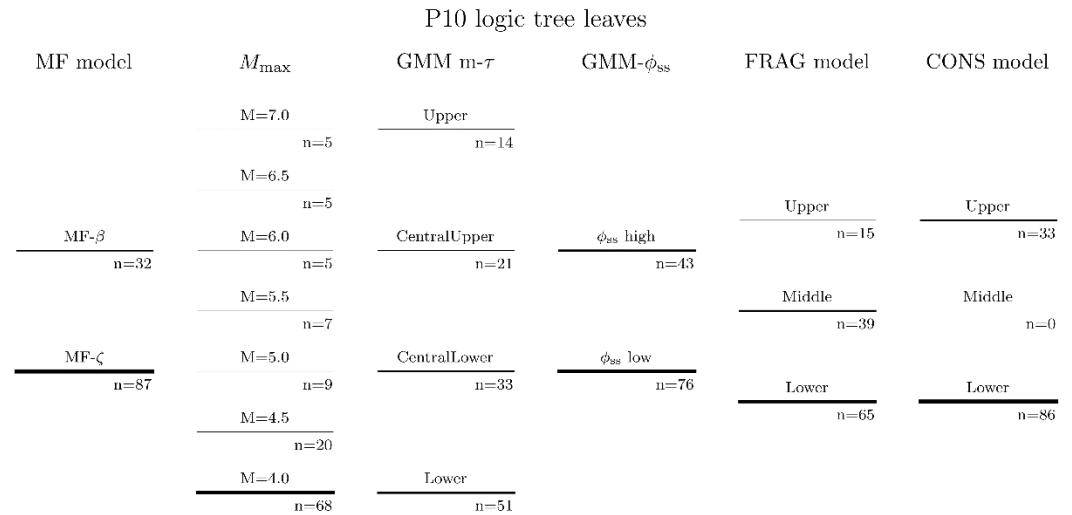


Figure 20: : Logic tree showing how the <P10 leaves are distributed over the logic tree branches. Number of leaves on each branch given by n . Line thickness of each branch proportional to n . See Figure 1 for logic tree weights.

3.7.2 LPR of individual buildings

The standard risk metric and the 1% risk metric are field-wide measures, therefore the change between test- and base case does not always reveal what the driver is for those risk variations. For many test cases, it is intuitively apparent where the risk variations come from (for instance, field-wide changes such as magnitude-frequency models, typology-specific changes, and alternative varying spatial data). If so desired, the risk drivers can be understood better by looking at the mean LPR per building, instead of the field-wide mean. We shall illustrate this for the risk variations between the base case and the alternative fault case. This test case was chosen because it contains spatial variability with respect to the faults in the base case. Note that the mean LPR over the entire building database of the alternative fault case is a factor 0.99 of that of the base case.

The alternative fault case reveals that risk increases for buildings that had a higher risk already in the base case (Figure 21, top left, $LPR > 1 \times 10^{-6}$). Buildings at a lower risk in the base case either have a lower or a higher risk in the alternative fault case. Normalised by the base case, we see that the building risks vary by between a factor 0.65 and 1.6 over the entire building database (Figure 21, top right). The variations in risk show a clear spatial trend (Figure 21, bottom), where the areas containing buildings with an increased risk also have an increased fault density. Hence, we observe plenty of variation in risk per individual building, as an effect of spatial variations in fault density, notwithstanding the nearly similar field-wide mean risks over the entire building database for both cases.

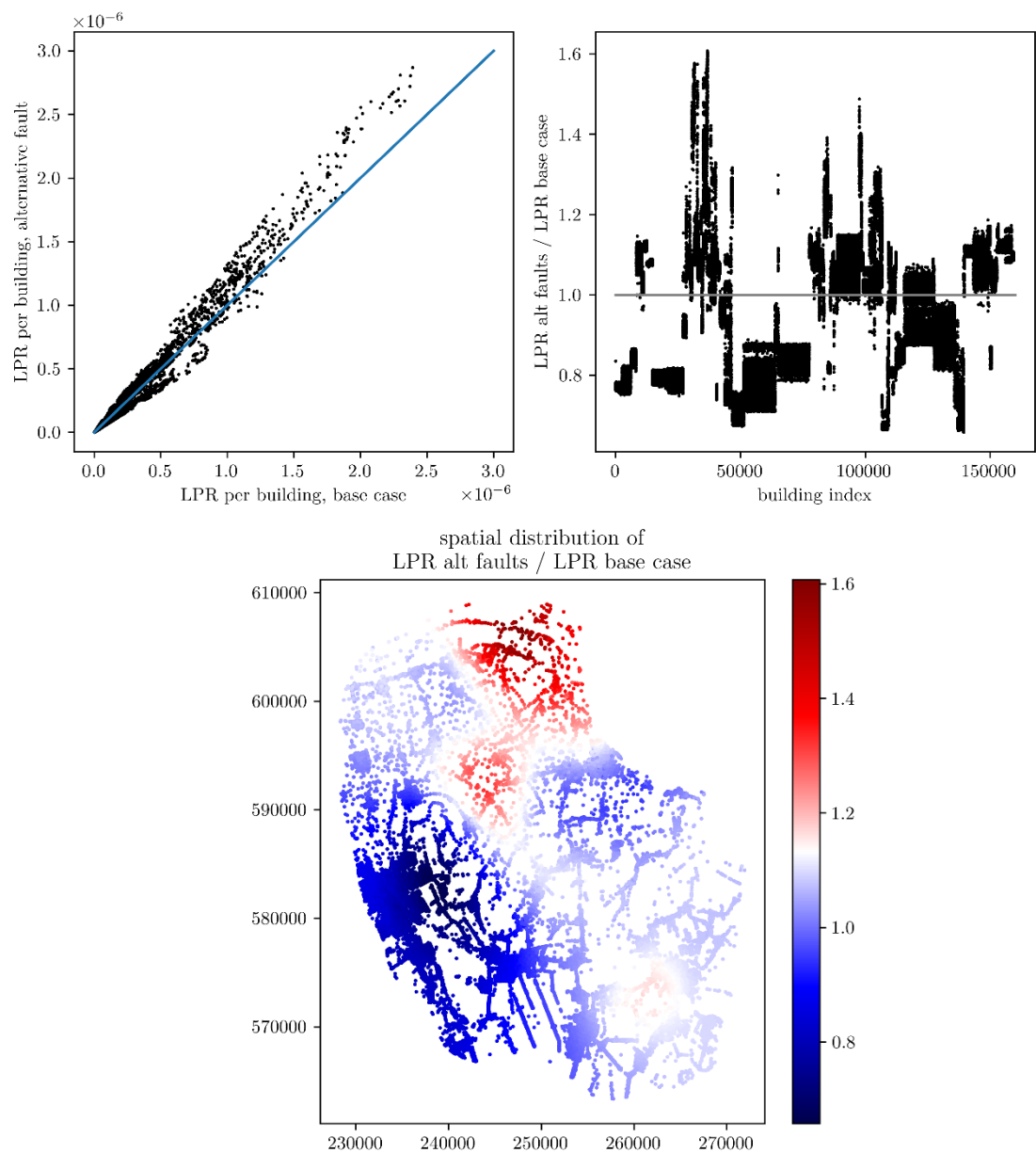


Figure 21: Top left: LPR per building for the base case versus LPR per building for the alternative fault case. Top right: LPR per building for the alternative fault case normalised by the base case for each building. Bottom: spatial distribution of the normalised LPR per building.

4 Discussion & conclusion

The sensitivity analyses that yield considerable changes in risk are summarised in Figure 22, together with the sensitivity to the six logic tree branches. The sensitivity range for $M_{\min}$ and calibration date range result from using a range of values. The sensitivity for Coulomb stress integration, period-period correlation, and magnitude-frequency distributions result from different model choices/implementations. The sensitivity for the parameters in the URM typology should be interpreted differently, as the four test cases each change a different model setting. However, the FCM sensitivity illustrates that mean risk is most sensitive to the fragility parameter of the URM typology. The magnitude-frequency model choices can be visualized as an extension to the existing MF branch in the logic tree.

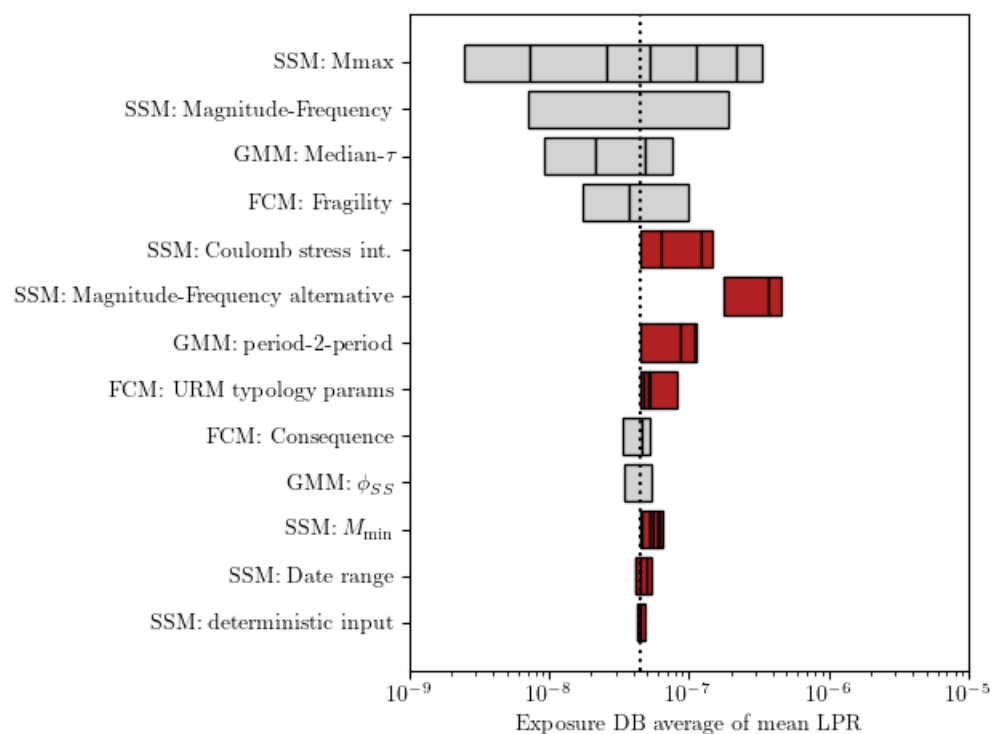


Figure 22: Extended tornado plot, showing the sensitivity of the mean risk for the logic tree branches (grey bars) and seven families of tested model settings (red bars).

The results presented here confirm the results from a sensitivity analysis performed by TNO as part of the pSHRA2021 [TNO, 2021a] to a first order, showing indeed that model choices in the GMM period-2-period correlation and Coulomb stress field integration, as well as alternative FCM parameters, have a significant impact on the simulated risk. We also stress the impact of the magnitude-frequency distribution model on the mean risk. The tested alternative MF models are near or exceed the highest risk MF model represented in the logic tree branch (the stress-dependent hyperbolic tangent model). Together with the already existing magnitude-frequency logic tree branch, the MF model choices give the largest range in predicted risks. We may also conclude that the predicted mean risk is not sensitive to most tested static deterministic data, and only mildly sensitive to compaction coefficient grid changes. However, local risk (i.e., mean risk of individual buildings) may change

from variations in deterministic data such as alternative faults, whilst the mean risk over all buildings remains constant.

The sensitivity to the inclusion or exclusion of ETAS is negligible. Last, we have quantified the sensitivity to the earthquake data used for calibration of the seismic source model. Relative minor changes in both date range and completeness change the mean risk by tens of percent.

We further note that the mean risk for nearly all test cases, and the base case, falls above the P90, indicating a highly skewed risk distribution. The relevance of the P90 percentile as a risk metric thus remains debatable.

Above, we summarised a variety of changes in mean risk for the tested case. Some very small changes may be neglected, whilst other test cases that show larger changes should be considered in the further development of the Model Chain Groningen. We propose a quantitative threshold in relative risk change, below which we consider the changes not relevant. This threshold is based on a number of test cases, most notably the choice of $M_{\min}$ or the calibration periods, that reflect aleatory variability in the input data. These cases involve (sub) selection of the dataset (catalogue) and should, in theory, for higher data volumes, not influence the result of the risk calculations. The fact that they do, may be interpreted as the result of random (aleatory) variations. These variations can only be reduced by incorporating more (denser) data, which is evidently not possible at any instance of time. Epistemic uncertainties that do not exceed aleatory variations may not require prioritization.

Of these test cases, the largest variation relative to the base case is found for $M_{\min}=1.7$. This change is a factor 1.4 with respect to the base case. Hence, we do not pay much attention to test runs with risks that change by a factor 0.7 up to 1.4. This leaves a smaller number of model settings with significant (i.e., exceeding aleatoric variation in input data) changes in risk (Figure 23), summarized in the following list:

- **GMM period-to-period correlation:** Full vs. absent period-period-correlation in the site response.
- **SSM Coulomb stress treatment for calibration:** Consistent approach for the Coulomb stress calculation and its use in the calibration of the seismic source model.
- **FCM parameters:** Fragility parameters for unreinforced masonry buildings.
- **SSM Frequency-Magnitude models:** The choice of a suitable magnitude-frequency model and a suitable predictor (covariate).

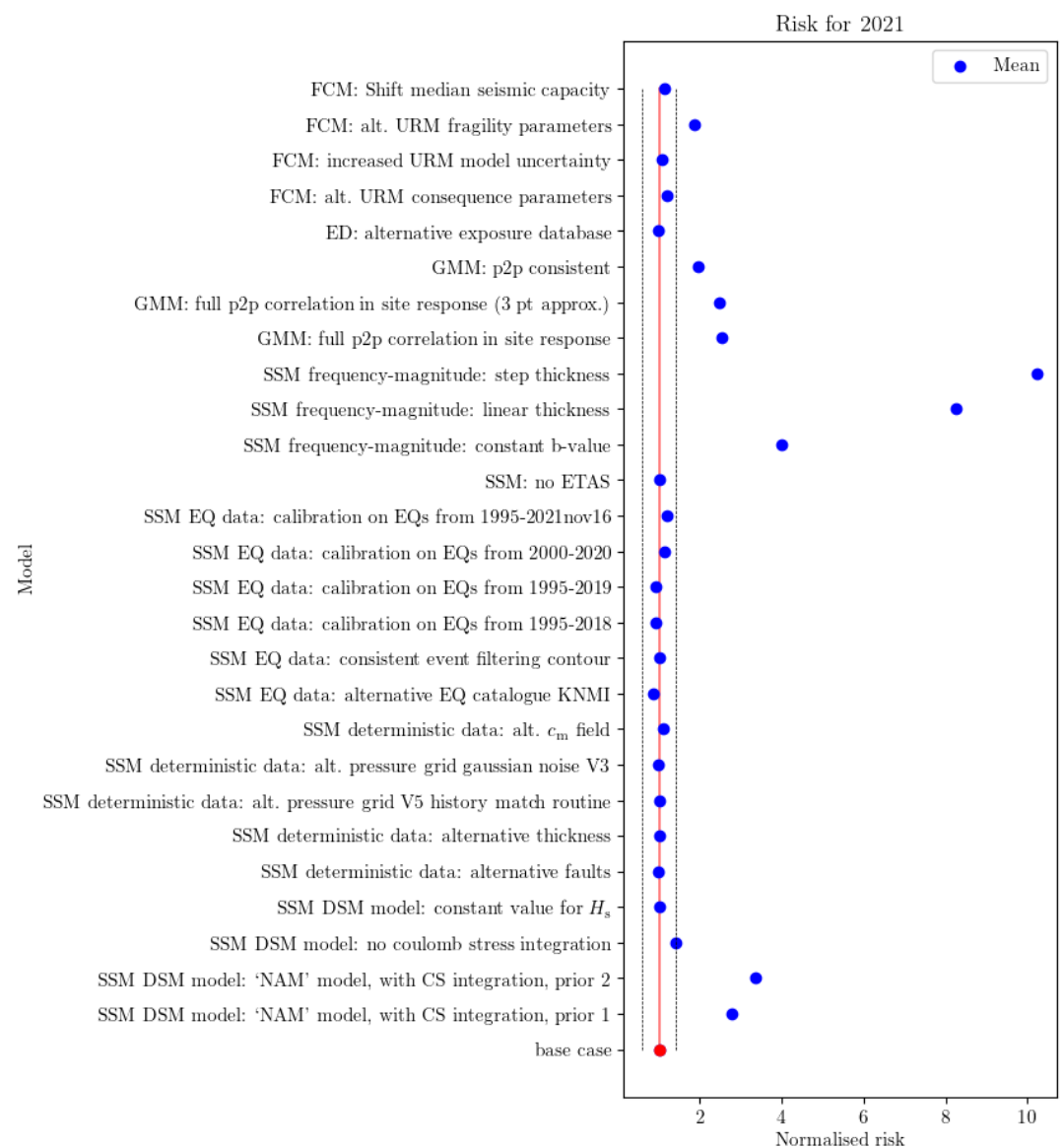


Figure 23: Risk for test cases normalised with respect to the base case. The black dotted lines highlight the factor 0.6 to 1.4 band, outside of which we deem risk changes significant.

5 References

Bommer, J., B. Edwards, P. Kruiver, A. Rodriguez-Marek, P. Stafford, M. Ntinalexis, E. Ruigrok, & B. Dost (2021). V7 Ground-Motion Model for Induced Seismicity in the Groningen Gas Field. 10 October 2021.

Bourne, S. J., & Oates, S. J. (2017). Extreme threshold failures within a heterogeneous elastic thin sheet and the spatial-temporal development of induced seismicity within the Groningen gas field. *Journal of Geophysical Research: Solid Earth*, 122(12), 10-299.

Bourne, S. J., Oates, S. J., & Van Elk, J. (2018). The exponential rise of induced seismicity with increasing stress levels in the Groningen gas field and its implications for controlling seismic risk. *Geophysical Journal International*, 213(3), 1693-1700.

Crowley, H., Pinho, R., & Cavalieri, F. (2019). Report on the v6 Fragility and Consequence Models for the Groningen Field, March 2019.

Dost, B., Goutbeek, F., van Eck, T., & Kraaijpoel, D. (2012). Monitoring induced seismicity in the North of the Netherlands: status report 2010. *KNMI Scientific report*, 2003.

Kortekaas, M., & Jaarsma, B. (2017). Improved definition of faults in the Groningen field using seismic attributes. *Netherlands Journal of Geosciences*, 96(5), S71-S85.

Kraaijpoel, D., Martins, J., Osinga, S., Vogelaar, B., & Breunese, J., (under review). Statistical analysis of static and dynamic predictors for seismic b-value variations in the Groningen gas field. submitted to: *Netherlands Journal of Geosciences*.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401), 9-27.

Varty, Z., Tawn, J. A., Atkinson, P. M., & Bierman, S. (2021). Inference for extreme earthquake magnitudes accounting for a time-varying measurement process. *arXiv preprint arXiv:2102.00884*.

NAM (2021). Groningen long term subsidence forecast. NAM 2021 EP202008201822.

TNO (2020a). Probabilistic Seismic Hazard and Risk Analysis in the TNO Model Chain Groningen, TNO 2020 R11052.

TNO (2020b). Status of the TNO Model Chain Groningen per October 1, 2020 and recommendations for the public Seismic Hazard and Risk Analysis 2021. TNO 2020 R11464, 9 October 2020.

TNO (2020c). TNO Model Chain Groningen: Update and quick scan comparison of 2020 HRA model, TNO 2020 R11659.

TNO (2020d). Advies vaststellingsbesluit Groningen gasveld 2020/2021, AGE 20-10.043, 11 mei 2020.

TNO (2020e). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen, TNO 2020 R10698B, 1 juni 2020.

TNO (2020f). Typologie-gebaseerde beoordeling van de veiligheid bij aardbevingen in Groningen – Addendum, TNO 2020 R11102, 15 september 2020.

TNO (2021a). Publieke Seismische Dreigings- en Risicoanalyse Groningen gasveld 2021. TNO2021 R10441, 24 maart 2021.

TNO (2021b). Status of the TNO Model Chain Groningen per October 1, 2021 and recommendations for the public Seismic Hazard and Risk Analysis 2022. TNO 2021 R11742, 15 October 2021.

Signature

Utrecht, 9 September 2022

TNO

J.A.J. Zegwaard
Head of department